The Combination of Different Modalities in the Web-Based Learning Environment: A Comparative Analysis of the Perceptual Instructional Outcomes

Parveen Kaur, Hanafi Atan, [+]Soon Fook Fong, Rozhan M Idrus & *Hisham Dzakiria
School of Distance Education, Universiti Sains Malaysia, 11800 Penang, Malaysia
Parveen_Kaur@toshibatec-tim.com.my, ahanafi@usm.my, rozhan@usm.my

[+] *School of Educational Studies, Universiti Sains Malaysia, 11800, Penang, Malaysia*
sffong@usm.my

**Faculty of Communication and Modern Languages, Universiti Utara Malaysia, 06010, Kedah, Malaysia*
hisham@uum.edu.my

Abstract

The Web-based Learning Environment (WBLE) has the capability to incorporate many modalities (text, graphics, audio and video) and their combinations into one single medium for instructional purposes. However, when each modality combination is used, information is extracted differently through the information pathways by the learners, resulting in varying levels of memory storage. This article reports on the comparative analysis between the three different types of modality combinations (text + graphics, text+ graphics + audio, text+ graphics +video) in the WBLE in terms of the instructional outcomes as perceived by the students. The Post Hoc Test for multiple comparisons revealed that the hierarchical modality combination with text + graphics + audio exhibited the highest mean score followed by the text + graphics combination. The text + graphics + video combination exhibited the lowest mean score. The effect of information degradation from the excessive rates of transmission through multi-channel communication is highlighted and discussed.

1. Introduction

The perceptual modalities refer to the primary ways the human body takes in information. The information is extracted from the surroundings by the human senses and the information passes through the information pathways and is retained in the memory. The perceptual modalities include text, graphics, audio and video and they are in the domain of the visual, verbal and non-verbal sensory systems.

The advancement of Information and Communication Technology (ICT) has a pronounced impact on the way these modalities and their combinations are being utilised in the Web-Based Learning Environment (WBLE) for instructional purposes. The WBLE has the ability to incorporate the usage of multi-modalities in a single medium that is capable of massive information flow to the learners. However, limitations may appear in the processing of such information by the human receptor. If not handled correctly, the multi-channel sensory information processes may be detrimental to the learning process, as multiple representations on the screen may place additional and unnecessary cognitive demands on a learner.

Several theoretical frameworks have been proposed to explain the use of multi-modalities for instruction [1]. Among these theoretical frameworks, the Cagne Information Processing Theory (CIPT) [2], the Dual Coding Theory (DCT) [3] and the Cue Summation Theory (CST) [4] have generated much interest. The Cagne Information Processing Theory describes how the human memory system acquires, elaborates, and retrieves information. The principal components of memory are the short-term and long-term memory. Knowledge stored in the long-term memory must be organised effectively for future recall. The structure of this knowledge reflects the manner in which the information is presented and perceived. The DCT, on the other hand, provides a strong explanation linking verbal and non-verbal subsystems. It assumes that there are two cognitive subsystems that are specialised for encoding, organising, transforming, storing and retrieving information [3]. These

Proceedings of the Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)
0-7695-2338-2/05 $20.00 © 2005 IEEE

subsystems are the non-verbal system which specialises in dealing with images of non-verbal objects and events and the verbal system which specialises in linguistic information. This theory predicts that the image system combines visual, auditory, and other sensory components of non-verbal information in integrated wholes [5]. Thus, an item that is presented in two modalities or more is independently represented in each encoding modality and the item can be recalled from either of those modalities.

The Cue Summation Theory predicts that learning increases as the number and types of cues increase [6]. It asserts that learning will be increased when stimuli of shared information are presented because they reinforce each other. This theory assumes that the synchronous presentation of information through different sensory channels would provide additional stimuli for reinforcement. Some of the recent studies conducted on the effect of computer-based multimedia as instructional means centre on the applicability of the Cue Summation Theory.

There has been considerable interest in looking at how modalities and their combinations perform in instructional outcomes and their relationship to the theories proposed. Bither & Wright [7] and Drew & Grimes [8] explicitly investigated effects of single versus dual modalities, and discovered that the recall of message contents was greater for audio-video messages. This result implies that two modalities are better than single modalities which support the Cue Summation Theory. Leigh [9], on the other hand, revealed that highly congruent audio and video stimuli exhibit comparable positive effects on information processing. Sundar et al., [10] explored the effects on the memory of two modalities (text-audio) versus single modalities (text only) within the computer mediated communication environment. They showed that there was a significant difference in memory retention between the two groups.

It is evident that there is a need to study the effects of various modality combinations on instructional outcomes in the WBLE. Accordingly, this study compares the impact of various combinations as perceived by the learner. The three combinations utilised were the text + graphics, text + graphics + audio and text + graphics + video. In conducting this study, the following research questions were raised:

- Is there a significant difference in terms of the knowledge enhancement as perceived by the students when exposed to the three different modality combinations?
- What is the hierarchy of the modality combinations in terms of the knowledge enhancement as perceived by the students when exposed to the three modality combinations?

The results of this study are important to assist in the provision of an appropriate instructional design in the course delivery for the WBLE that involves a varied combination of perceptual modalities. This would ensure an effective information flow for memory processing which results in an enhancement of instructional performance, making learning enjoyable and meaningful.

2. Methodology

For the purpose of this study, three different types of WBLEs were developed. They were identical in terms of the instructional design used which closely followed the Allessi & Trollip [11] instructional design model. The contents of the Web pages were also identical, consisting of the topic *The Atomic Bohr Model,* based on the curriculum of the first year ZCT 104 - Modern Physics course offered by the School of Physics, Univeristi Sains Malaysia, Malaysia. The independent variables were the three types of modality combinations used, namely, text + graphics, text + graphics + audio and text + graphics + video. The dependent variable was the perceived knowledge enhancement derived from the difference of the post- and the pre- formative mean scores.

The sample consisted of 169 students randomly selected from a total of more than 400 students registered for the course. They were found to be homogenous in terms of the background knowledge related to the contents of the WBLE as well as regarding computer competency. The samples were assigned according to the treatment groups, with 40, 75 and 54 students in the text + graphics, text + graphics + audio and text + graphics + video groups respectively.

The research procedure involved administering the 4-Likert scale pre-formative questionnaire consisting of 7 items related to the course contents to the students prior to the treatment. The students were required to respond and indicate their level of understanding of the course contents of the WBLE. They subsequently underwent an hour's exposure to their corresponding WBLE, and where necessary by virtue of the treatment group, they were required to download sound or video streaming for instructional purposes. Post-formative questionnaires were administered immediately after the treatment to gauge the level of their newly constructed knowledge as well as the extent of learning that had taken place. The analysis of data was carried out using the standard SPSS version 10.0 statistical packages.

3. Results and discussion

Table 1 depicts the ANOVA analysis of the pre-formative mean scores of the three treatment groups.

The analysis yielded a value of F=1.188 at p=0.368 significance level, indicating that there were no significant differences between the groups in terms of their background knowledge prior to the treatment. Irrespective of the type of modality treatment received by the samples, all the respondents in this study were essentially the same on the measure of the dependent variables at the beginning of the study. Students in the three treatment groups were homogenous in their perceptual background knowledge before they received the treatment. According to Sabrina (1997), it is imperative that the samples are homogenous prior to the treatment so that any shift in the range of the post-formative score would result from the treatment provided with no contamination from the sample's prior knowledge of the treatment contents.

Table 1. ANOVA of pre- and post-formative mean scores of treatment groups.

Formative	Treatment Groups	Sum of squares	df	Mean Square	F	Sig
Pre-	Btwn groups	1.020	2	0.510	1.188	0.368
	Within groups	69.787	164	0.425		
Post-	Btwn groups	17.541	2	8.770	12.736	0.000*
	Within groups	113.204	164	0.684		

The mean difference is significant at the 0.05 level (p*<0.05)

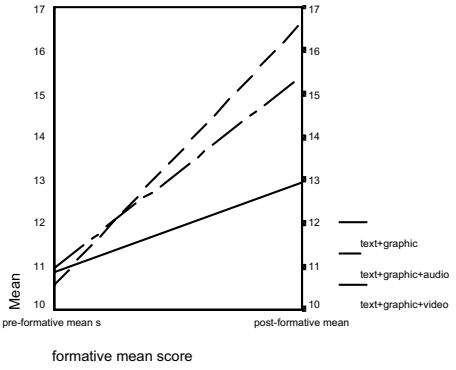

Figure 1. Knowledge enhancement as perceived by the treatment groups.

Figure 1 depicts the descriptive analysis of the pre-formative and post-formative mean scores. It is evident that the attainment of low pre-formative mean scores compared to the post-formative mean scores was an indication of non-mastery on the treatment content before the treatment. The increase of the mean scores after the treatment is indicative of the perceived knowledge enhancement experienced by the students. The figure also illustrates that the text + graphics + audio treatment group recorded the highest mean gain, followed by the text + graphics treatment group. The text + graphics + video treatment group revealed the lowest mean gain.

The ANOVA analysis of the post-formative mean score is also shown in Table 1 with the value of F=12.736, p=0.00. This result indicates that there was a significant difference in the post-formative mean scores among the three different modality groups. To elucidate the hierarchy among the three modality combinations, a Post Hoc Multiple Comparison Analysis was carried out.

Table 2 illustrates the Post Hoc Multiple Comparisons Analysis of the mean gain scores of the three treatment groups. The significant differences arose from the comparison between the text + graphics modality combination with that combining the text + graphics + video (p=0.025) as well as between the text + graphics + audio combination with that combining the text + graphics + video (p=0.000). The comparison between the text + graphics and text + graphics + audio combinations exhibited no significant differences (p=0.36). The Post Hoc Multiple Comparisons Analysis and the post-formative mean scores as depicted in Figure 1 demonstrate that the highest performances in terms of the students' perceived instructional outcomes were achieved by the text + graphics + audio modality combination followed by the text + graphics combination. The text + graphics + video combination indicated the least satisfactory performance.

Table 2. Post Hoc Multiple Comparisons Analysis of post-formative scores

Treatment group (I)	Treatment Group (J)	I-J	Std. Error	Sig
text + graphics + audio	text + graphics + video	0.75	0.150	0.000*
	text + graphics	0.26	0.163	0.306
text + graphics + video	text + graphics + audio	-0.75	0.150	0.000*
	text + graphics	-0.49	0.175	0.025*
text+ graphics	text + graphics + audio	-0.26	0.163	0.306
	text + graphics + video	0.49	0.175	0.025*

Taking into consideration all the three modality combinations from the above analysis, a hierarchy of the types of modalities can as constructed as shown in Figure 2.

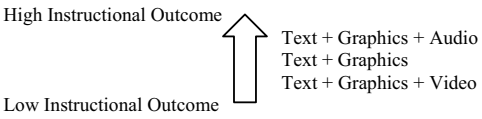

Figure 2. Hierarchy of the types of modality combinations.

The results revealed that the students perceived the highest knowledge enhancement in the instructional process that incorporated the text + graphics + audio modality combination in the WBLE as compared to the text + graphics combination alone. The text + graphics combination with the audio narration allowed the students to hear and understand certain announcements of formulas and symbols involved and relate the narration to the accompanied text. Such a modality combination presents relevant auditory and visual cues that are clear, concise and easily detectable and provide sufficient cognitive capacity to enable students to integrate them into their working memory.

This finding is consistent with the study undertaken by Sundar et al. [10]. They revealed that students who were exposed simultaneously to both textual and auditory modalities obtained better memory gain compared to students who read the text only. The students using the audio streaming had better control over their learning. Typically, when people listen to an audio download on the computer, they can actively chose to either forward or rewind their material to reinforce understanding instead of passively listening to the audio streaming.

It is also evident that the text + graphics + video modality combination was perceived the least satisfactory by the students in terms of the immediate recall ability of the knowledge enhancement. The possible explanation to this is that the unfamiliarity with the combination of video + sound download might have distracted the students from their concentration on the course contents being conveyed. As opposed to the audio streaming, the video streaming created a new video page which overlapped with the screen text. The disappearance of the screen text resulted in the students being unable to correlate the narration with the text, hence losing concentration on the course contents.

Another plausible explanation is the degradation of the information processing when more modalities are added. In contrast to the Cue Summation Theory, the findings in this study show that the addition of the extra modalities increased the cognitive load and this had the potential to lead to a less efficient processing of information in the brain. Broadbent [12] explained that it is possible to have a concentration loss from excessive rates of transmission with multi-channel communication because the message recipient would normally find it difficult to alternate successfully between different modality channels. The multi-channel communication presented additional cues that reduced the cognitive capacity used to reinforce the working memory. The substantial reduction of cognitive capacity slowed down the integration of the additional visual and the auditory cues in the students' mental representation leading to lower immediate recall ability as observed in this finding

4. Summary

This study shows that different modality combinations in the WBLE have an effect on the instructional outcomes as perceived by the students. Even though the results obtained are influenced by research limitations such as the course involved as well as the instructional design used, the results have successfully highlighted the importance of the selection of an appropriate mix of modalities when designing course contents in the WBLE. Appropriate modality combinations must be suitably provided to ensure an effective flow of information that would result in the enhancement of instructional outcomes.

5. References

[1] Burton, J.K., Moore, D.M. & Holmes, G.A. (1995). Hypermedia concepts and research: an overview. *Computers in Human Behavior*, 11(3/4), 345-369

[2] Cagne, R.M. (1985). *The Conditions of Learning and Theory of Instruction.* Fort Worth: Holt, Rinehart and Winston Inc.

[3] Pivio, A. (1986). *Mental Representation*. New York: Oxford University Press.

[4] Severin, W. (1967). Another look at the Cue Summation. *Audiovisual Communication*, 15(3), 233-245

[5] Paivio, A. & Begg, I. (1981). *The Psychology of Language.* New York: Prentice Hall

[6] Moore, D.M., Burton, J.K. & Myers, R.J. (1996). Multiple-channel communications: The theoretical and research foundations of multimedia. In D.H. Jonassen (Ed.). *Handbook of Research for Educational Communications and Technology.*
http://www.aect.org/Intranet/Publications/edtech/29/index.html

[7] Bither, S.W. & Wright, P.L. (1973). The self-confidence-advertising response relationship: A function of situational distraction. *Journal of Marketing Research*, 10, 146-152.

[8] Drew, D.G. & Grimes, T. (1987). Audio-visual redundancy and TV news recall. *Communication Research*, 14, 452-461.

[9] Leigh, J.H. (1992). Modality congruence, differences among broadcast media: review and suggestions. *Journal of Advertising*, 20, 71-75.

[10] Sundar, S.S., Kalyanaraman, S. & Jones, E. (2000). Modality effects on memory for multimedia messages. Paper presented at the International Association for media and Communication research, Singapore, July.

[11] Alessi, S.M. & Trollip, (2001). *Multimedia for Learning and Development*. Massachussetts: Allyn & Bacon.

[12] Broadbent (1958). *Perception and Communication*. New York: Pergamon Press.